Validation of Aura Microwave Limb Sounder O_3 and CO observations in the upper troposphere and lower stratosphere

N. J. Livesey[1], M. J. Filipiak[2], L. Froidevaux[1], W. G. Read[1], A. Lambert[1], M. L. Santee[1], J. H. Jiang[1], H. C. Pumphrey[2], J. W. Waters[1], R. E. Cofield[1], D. T. Cuddy[1], W. H. Daffer[1], B. J. Drouin[1], R. A. Fuller[1], R. F. Jarnot[1], Y. B. Jiang[1] B. W. Knosp[1], Q. B. Li[1], V. S. Perun[1], M. J. Schwartz[1], W. V. Snyder[1], P. C. Stek[1], R. P. Thurstans[1], P. A. Wagner[1], M. Avery[3], E. V. Browell[3], J.-P. Cammas[4], L. E. Christensen[1], G. S. Diskin[3], R.-S. Gao[5], H.-J. Jost[6], M. Loewenstein[7], J. D. Lopez[7], P. Nedelec[4], G. B. Osterman[1], G. W. Sachse[3], C. R. Webster[1]

Abstract. Global satellite observations of ozone and carbon monoxide from the Microwave Limb Sounder (MLS) on the EOS Aura spacecraft are discussed with emphasis on those observations in the 215 – 100 hPa region (the upper troposphere and lower stratosphere). The precision, resolution and accuracy of the data produced by the MLS 'version 2.2' processing algorithms are discussed and quantified. O_3 accuracy is estimated at ∼40 ppbv+5% (∼20 ppbv+20% at 215 hPa) while the CO accuracy is estimated at ∼30 ppbv+30% for pressures of 147 hPa and less. Comparisons with expectations, model results and other observations show good agreements for the O_3 product, generally consistent with the systematic errors quoted above. In the case of CO, a persistent factor of ∼2 high bias is seen at 215 hPa, however the morphology is shown to be realistic, consistent with raw MLS radiance data, and useful for scientific study. The MLS CO data at higher altitudes are shown to be consistent with other observations and model results.

1. Introduction

Ozone and carbon monoxide play important and distinct roles in the upper troposphere. Upper tropospheric ozone is a potent and poorly understood greenhouse gas [*Intergovernmental Panel on Climate Change*, 2001] whose abundance, ranging from tens to a few hundred parts per billion (ppbv), is influenced by a variety of factors. These include the abundance of precursor HO_x and NO_x species and influx of ozone rich air from the lower stratosphere. Rapid transport of air from the boundary layer to the upper troposphere by deep convection strongly affects ozone through transport of precursor species [*Prather and Jacob*, 1997]. Carbon monoxide is a byproduct of combustion, both natural and anthropogenic, and is the primary sink of tropospheric OH, the main atmospheric oxidant. Its relatively long (∼2 month) photochemical lifetime in the upper troposphere makes it useful as an indicator of other pollutants such as carbonaceous aerosols, CH_4 and other hydrocarbons. CO is also useful as a tracer of atmospheric motions, particularly of the long-range transport of polluted air [*Stohl et al.*, 2002; *Liu et al.*, 2003].

The Microwave Limb Sounder (MLS) [*Waters et al.*, 2006] on the Aura spacecraft [*Schoeberl et al.*, 2006b], launched on 15 July 2004, observes thermal microwave limb emission from many molecules, including O_3 and CO. This paper describes MLS O_3 and CO data in the upper troposphere and lower stratosphere (UT/LS), which is broadly defined here as the region from ∼300 – 100 hPa. The data described are those produced by version 2.2 of the MLS data

[1]Jet Propulsion Laboratory, California Institute of Technology
[2]University of Edinburgh
[3]NASA Langley Research Center
[4]Laboratoire d'Aérologie, UMR5560 CNRS, University P. Sabatier, Toulouse
[5]NOAA Earth Systems Research Laboratory, Chemical Sciences Division
[6]Novawave Technologies, 900 Island Dr., Redwood City, California
[7]NASA Ames Research Center

processing algorithms. The precision, accuracy and resolution of these data are discussed, both through the expected impacts of uncertain knowledge of instrument calibration etc., and through comparisons with expectations and other observations of these species.

Validation of the CO observations at higher altitudes is discussed by *Pumphrey et al.* [2007]. The MLS ozone product is also described in two companion papers in this issue. *Froidevaux et al.* [2007] focus on observations in the stratosphere and mesosphere, while *Jiang et al.* [2007] describe comparisons of MLS O_3 data with sonde and ground-based observations, including in the altitude region discussed in more detail in this paper.

Section 2 describes the relevant aspects of the MLS instrument and data processing strategy, gives rules on appropriate screening for the UT/LS O_3 and CO data, and quantifies their typical precision, expected accuracy and spatial resolution. Section 3 describes some 'zero order' validation of these data by comparisons with expectations, non-coincident observations and model calculations. Section 4 focuses on comparisons between MLS data and various co-located aircraft based observations of O_3 and CO. Finally, section 5 summarizes all these findings, reports on remaining issues with the validation of these MLS data, and outlines plans for further validation and future versions of the products.

2. MLS UT/LS O_3 and CO observations

2.1. MLS instrument operations and data description

MLS observes thermal microwave emission from the Earth's limb in five spectral regions from 118 GHz to 2.5 THz. The O_3 and CO standard products described in this paper are taken from observations in the 230 – 250 GHz spectral range. MLS looks forward from the Aura spacecraft and scans the Earth's limb vertically from the ground to ∼90 km every 24.7 s. The vertical scan rate varies with altitude, with a slower scan (giving a better signal to noise through greater integration time) used in the lower regions (∼0 – 25 km). The MLS vertical scans are synchronized to the Aura orbit such that vertical scans are made at essentially the same latitudes each orbit, with 240 scans performed per orbit (∼3500 scans per day).

This paper describes MLS 'Level 2' data, which are geophysical products reported along the measurement track of the instrument. These are retrieved from calibrated MLS radiance observations ('Level 1 data') by the MLS data processing software [*Livesey et al.*, 2006]. The MLS Level 2 products are reported on a fixed vertical pressure grid having 6 levels per decade change in pressure in the troposphere and stratosphere, evenly spaced in $\log_{10}$ pressure starting at 1000 hPa (thinning out to 3 per decade at pressures less than 0.1 hPa). These profiles are evenly spaced at 1.5° great circle angle (geodetic) along the orbit track. This gives 240 Level 2 profiles per orbit at fixed latitudes, synchronized to the MLS vertical scans.

The MLS Level 2 products are reported in Level 2 Geophysical Product (L2GP) data files. Individual files describe one MLS 'standard product' (O_3, CO, H_2O etc.) for a 24 hour period from midnight to midnight universal time. The L2GP files store the data in an HDF-EOS version 5 'swath' format with the swath name describing the product. The ozone files contain additional swaths giving the estimated column ozone amount above the tropopause [*Froidevaux et al.*, 2007]. The MLS Version 2.2 data quality document [*Livesey et al.*, 2007] gives more information on the format and contents of the MLS data files.

2.2. Proper use of MLS UT/LS O_3 and CO data

In addition to describing file formats and contents, the data quality document [*Livesey et al.*, 2007] also gives detailed instructions on the proper use of all MLS data products. The pertinent information for MLS UT/LS CO and O_3 is repeated here.

Each MLS Level 2 data point is reported with a corresponding precision value. These quantify the impact on the data of noise on the MLS radiance measurements and, particularly in regions of lower measurement sensitivity, the contribution of a priori information. These issues are discussed in more detail in section 2.4. As an aid to users, the precisions are set to negative values in situations where more than 25% of the information came from the a priori value and the product should not be used in scientific studies.

Three additional data quality metrics are provided in the data files for each vertical profile. 'Status' is an integer bit field indicating circumstances where profiles are not to be used, or may be suspect due to instrumental and/or retrieval issues. The interpretation of the various bits in 'Status' are given in Table 1. Odd values of status indicate profiles that should not be used. Non-zero, even values of 'Status' indicate situations where care may be needed. In most cases this indicates that the retrieval algorithm detected strong cloud signatures in some radiances and chose to discard individual radiance measurements. The impact of this on MLS data varies with species and height. In general such profiles are suitable for scientific use, though they are usually reported with poorer precision due to the fewer number of radiances used in their retrieval (note that this is a change from v1.5, where such profiles were to be ignored in the UT/LS). More details on the 'Status' field are given in the data quality document. The 'Quality' field gives a measure of the fit achieved

Table 1. Meaning of bits in the 'Status' field.

Bit	Value[a]	Meaning
0	1	Flag: Do not use this profile (see bits 8–9 for details)
1	2	Flag: This profile is 'suspect' (see bits 4–6 for details)
2	4	Unused
3	8	Unused
4	16	Information: This profile may have been affected by high altitude clouds
5	32	Information: This profile may have been affected by low altitude clouds
6	64	Information: This profile did not use GEOS-5 temperature a priori data
7	128	Unused
8	256	Information: Retrieval diverged or too few radiances available for retrieval
9	512	Information: The task retrieving data for this profile crashed (typically a computer failure)

[a] 'Status' field in L2GP file is total of appropriate entries in this column.

to the measured MLS radiances by the retrieval (larger numbers imply better fits). Finally, 'Convergence' compares the fit achieved for a 'chunk' of ∼10 profiles to that expected by the retrieval algorithms. Values in the range 1.0 – 1.1 indicate that good convergence has been achieved, larger values imply poorer convergence.

All of these data quality metrics need to be considered when using MLS UT/LS O_3 and CO data. For MLS CO and O_3 observations in the range 215 hPa to 100 hPa, data should only be used when:

1. The precision value for that data point is positive
2. The 'Status' field for that profile is an even number (note that this is different from the rules for the earlier version 1.5 data)
3. The 'Quality' field for that profile is greater than 1.2 (note this is stricter than the 0.4 threshold to be used for stratospheric and mesospheric O_3 and 0.2 for stratospheric and mesospheric CO)
4. The 'Convergence' field for that profile is less than 1.8 (note that the threshold for stratospheric and mesospheric O_3 is 2.0

The MLS O_3 and CO data are retrieved over the range 316 –0.00046 hPa. As will be shown later in this paper, the version 2.2 O_3 and CO data at 316 hPa are not considered useful for scientific study, and only data in the range 215 – 0.0022 hPa and lesser pressures should be used (see *Froidevaux et al.* [2007] and *Pumphrey et al.* [2007] for discssion of upper altitude limits).

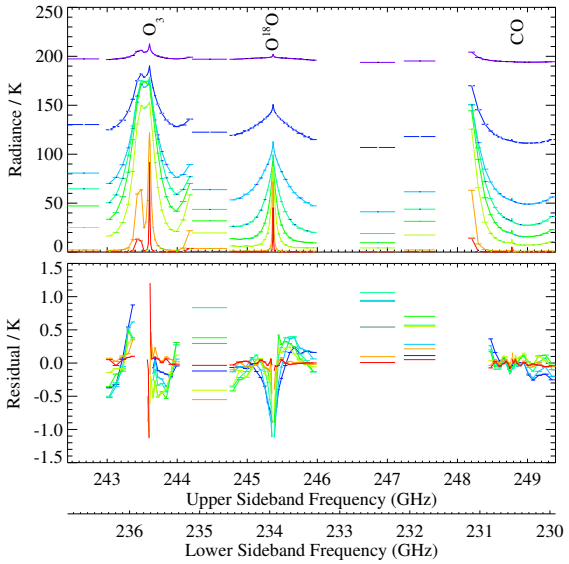

Figure 1. (Top) sample radiances (shown in terms of brightness temperature) from the MLS 240 GHz radiometer. Global average radiances from observations on 24 September 2004 are shown for 8 selected tangent point altitudes ranging from ∼7.5 km (purple) to ∼45 km (red). The MLS signal is a combination of incoming radiance at frequencies above (upper sideband, upper x-axis) and below (lower sideband, lower x-axis) the 239.660 GHz local oscillator. The widths of the various MLS spectral channels are denoted by the horizontal bars. The position of O_3, $O^{18}O$ and CO emission lines are noted. (Bottom) The average fit achieved to these radiances by the MLS version 2.2 retrieval algorithms. Some channels are not used in the retrievals, and so are not in the lower panel (e.g., those around 243.5 GHz upper sideband frequency).

2.3. Signature of UT/LS O_3 and CO in MLS radiances

Figure 1 shows typical MLS radiance observations in the 240 GHz region of the spectrum, from which the UT/LS O_3 and CO products are derived. All of the strong spectral features are due to emission from O_3 lines, with the exception of the feature at ∼234.0 GHz in the lower sideband, which is due to $O^{18}O$ emission (MLS band 8). The CO spectral line is at ∼230.5 GHz in the lower sideband and has a ∼1 K typical amplitude in the upper troposphere. The small features in this region seen at the higher tangent altitudes (e.g., red line) are due to strong emission from mesospheric CO. The UT/LS CO information derives from the radiances in this spectral region (MLS band 9). The UT/LS O_3 information derives from the broad spectral contrast across the band, mainly seen by three of the four 'wide' channels (MLS band 33), namely those at ∼244.5 GHz, ∼246.8 GHz and ∼247.5 GHz upper sideband frequency.

The MLS CO and O_3 data are retrieved using an optimal estimation approach [*Rodgers*, 2000; *Livesey et al.*, 2006] from these 240 GHz radiances, along with observations of the 118 GHz oxygen line for additional pointing information. In addition to O_3 and CO, this retrieval also produces estimates for HNO_3, SO_2, temperature, geopotential height and tangent pressure (along with spectrally flat 'extinction' terms). The lower part of Figure 1 shows the average fit achieved to these measured radiances by the retrieval algorithms. The scatter about these averages (not shown) is generally consistent with the levels of noise seen in the radiances, as would be expected and desired. The fits in the CO region are generally within ∼0.2 K, while the broad spectral structure, away from strong stratospheric features, is generally fitted within a few tenths of a Kelvin.

Although MLS observations are unaffected by thin cirrus clouds or stratospheric aerosols, thick clouds associated with deep convection can have an impact on the MLS radiances. The effect of emission and scattering from high altitude (≤∼200 hPa) clouds is to enhance the MLS radiance signals, while scattering by low clouds leads to suppression of observed limb radiances. These signatures are generally fairly spectrally flat. However, large amounts of scattering from the thickest clouds can attenuate the spectral variations in MLS radiances on which the composition measurements are based. The MLS data processing algorithms retrieve a spectrally flat 'extinction' term to compensate for scattering by moderate clouds. When the algorithms detect particularly thick clouds (through comparison of the observed radiances with predictions from a clear-sky-only radiative transfer model) that may significantly affect the spectral contrast, radiances from individual 1/6 s integration periods may be omitted from the retrieval. As discussed, this can lead to poorer precision being reported for the retrieved profiles.

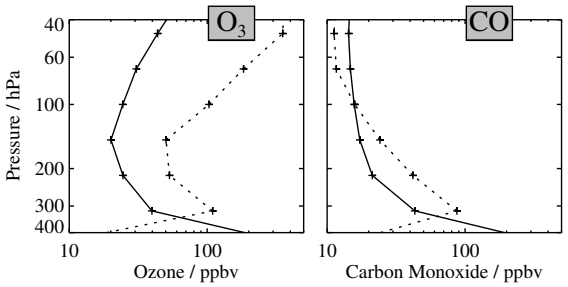

Figure 2. RMS estimated precision (solid line) and 1-σ scatter (dashed line) seen in the MLS UT/LS O_3 (left) and CO (right) measurements on 17 September 2004 in the 30°S – 30°N region.

2.4. Precision, scatter and spatial resolution

Each point in the retrieved MLS composition profiles is accompanied by an estimated 'precision' field. This is taken from the diagonal elements of the solution covariance matrix [*Livesey et al.*, 2006] and mainly reflects the contributions of radiance noise to the MLS measurements. In regions where MLS is less sensitive, the uncertainty on the a priori values used as virtual measurements begins to dominate the reported precision. Figure 2 summarizes the reported precision seen in MLS UT/LS O_3 and CO measurements on 17 September 2004 in the 30°S to 30°N region, typical of all these data (although the reported UT/LS O_3 precision in winter polar regions is ∼20–40% poorer). The root mean square (RMS) average of the estimated precision (solid lines) for O_3 in the UT/LS is 20 – 40 ppbv, with ∼15 – 40 ppbv estimated for CO.

It is useful to compare these precision estimates to the actual scatter seen in MLS data (broken lines). In cases where atmospheric variability is expected to be low compared to the MLS precision, this scatter will be comparable to the estimated precision (typically a little less, due to the influence of smoothing on the MLS retrievals [*Froidevaux et al.*, 2006]), as is seen here for CO in the mid-stratosphere. For the UT/LS CO and O_3 observations, the scatter is generally larger than the estimated precision, implying significant atmospheric variability and/or contributions from other sources of random error than radiance noise.

The MLS retrieval algorithms operate in a two dimensional 'tomographic' manner [*Livesey and Read*, 2000; *Livesey et al.*, 2006]. This approach allows for the direct modeling of the impact of gradients along the forward-looking line of sight of MLS. As with most remote sounding measurements, the resolution of the retrieved data can be describing using 'Averaging Kernels' [*Rodgers*, 2000]. The two-dimensional nature of the MLS retrieval system means that these kernels

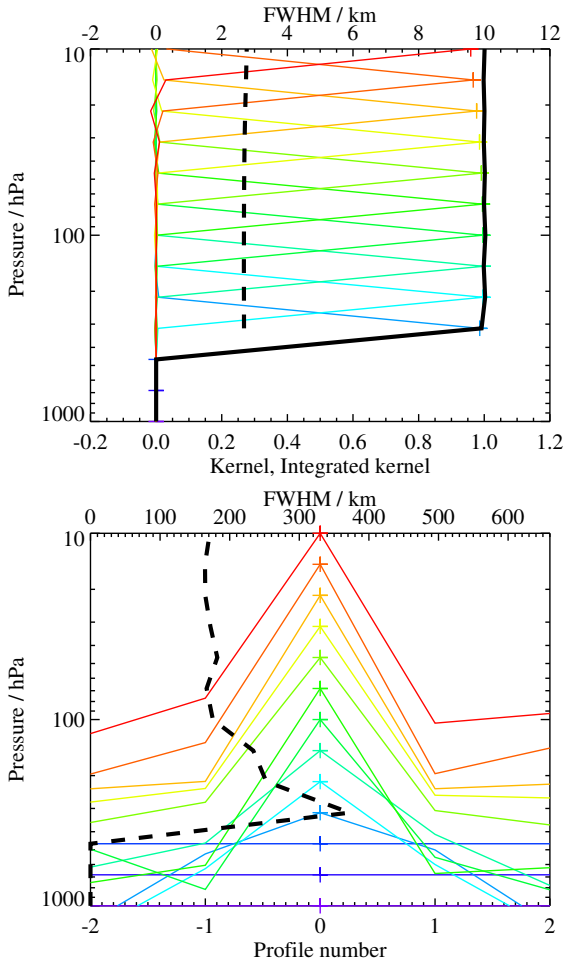

Figure 3. Typical two-dimensional (vertical and horizontal along-track) averaging kernels for the MLS v2.2 O_3 data at the equator; variation in the averaging kernels is sufficiently small that these are representative for all profiles. Colored lines show the averaging kernels as a function of MLS retrieval level, indicating the region of the atmosphere from which information is contributing to the measurements on the individual retrieval surfaces, which are denoted by plus signs in corresponding colors. The dashed black line indicates the resolution, determined from the full width at half maximum (FWHM) of the averaging kernels, approximately scaled into kilometers (top axes). (Top) Vertical averaging kernels (integrated in the horizontal dimension for five along-track profiles) and resolution. The solid black line shows the integrated area under each kernel (horizontally and vertically); values near unity imply that the majority of information for that MLS data point has come from the measurements, whereas lower values imply substantial contributions from a priori information. (Bottom) Horizontal averaging kernels (integrated in the vertical dimension) and resolution. The averaging kernels are scaled such that a unit change is equivalent to one decade in pressure.

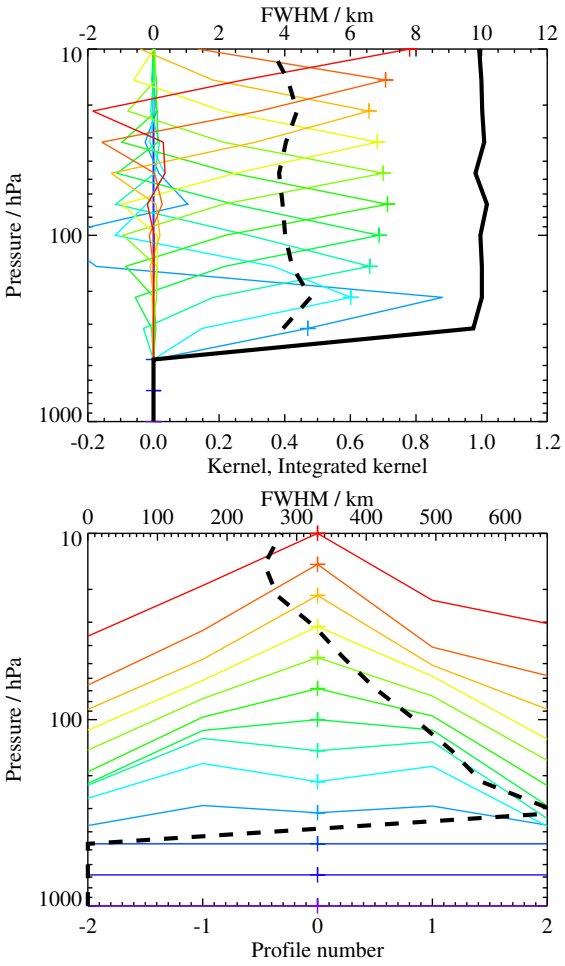

Figure 4. As Figure 3 but for MLS v2.2 CO.

describe both vertical and horizontal resolution. Figures 3 and 4 show horizontal and vertical aspects of the averaging kernels for UT/LS O_3 and CO, respectively. The vertical resolution of UT/LS O_3, as defined by the width of the kernels, is ∼2.5 km (essentially the same as the vertical spacing of the retrieval surfaces), while the CO data have poorer ∼4 km vertical resolution. For example, the 215 hPa MLS CO values derive ∼35% of their information from the atmospheric state at 147 hPa. The 316 hPa CO kernel has an unusual shape, indicating that these retrievals are more sensitive to CO at 215 hPa than 316 hPa, and show anti-correlations with CO at higher altitudes. In the along-track horizontal direction, the ozone data have single-profile resolution (∼165 km) at pressures of 100 hPa and less, with resolution closer to ∼350 km at greater pressures. The horizontal resolution of the CO observations is ∼500 – 600 km. The cross-track horizontal resolution for both products is defined by the horizontal width of the MLS field of view, which, for the 240 GHz radiometer that measures O_3 and CO, is ∼6 km.

2.5. Quantification of systematic uncertainties

2.5.1. Approach A major component of the validation of MLS data is the quantification of the various sources of systematic uncertainty. These can arise from instrumental issues (e.g., radiometric calibration, field of view characterization), spectroscopic uncertainty, and through approximations in the retrieval formulation and implementation. This section summarizes the relevant results of a comprehensive quantification of these uncertainties, performed for all MLS products. More information on this assessment is given in Appendix A of *Read et al.* [2007].

For each identified source of systematic uncertainty, its impact on MLS measurements of radiance (or pointing where appropriate) has been quantified and modeled. These modeled impacts correspond to either 2-σ estimates of uncertainties in the relevant parameter(s), or an estimate of their maximum reasonable error(s) based on instrument knowledge and/or design requirements. The impact of these perturbations on retrieved MLS products has been quantified for each uncertainty source by one of two methods.

In the first method sets of modeled errors corresponding to the possible magnitude of each uncertainty have been applied to simulated MLS cloud-free radiances, based on a model atmosphere, for a whole day of MLS observations. These sets of perturbed radiances have then been run through the routine MLS data processing algorithms, and the comparison between these runs and the results of the 'unperturbed' run used to quantify the systematic uncertainty in each case. The impact of the perturbations varies from product to product and among uncertainty sources. In some cases, the perturbation leads mainly to an additive bias in the product; in others, some multiplicative bias may be introduced. In most cases, some additional scatter is also introduced into the data.

Although the term 'systematic uncertainty' is often associated with consistent biases and/or scaling errors, many sources of 'systematic' error in the MLS measurement system give rise to additional scatter. For example, an error in the O_3 spectroscopy, while being a bias on the fundamental parameter, will have an impact on the retrievals of species with weaker signals (e.g., CO) that is dependent on the amount and morphology of atmospheric ozone. The extent to which such terms can be expected to average down is estimated to first order by these 'full up studies' through their separate consideration of the bias and scatter each uncertainty source introduces.

The difference between the retrieved product in the unperturbed run and the original 'true' model atmosphere is taken as a measure of uncertainties due to retrieval formulation and numerics. Another retrieval of the unperturbed radiances is performed with modifications made to the a priori information to test sensitivity to these constraints. In the case of O_3, the a priori value is increased by 200 ppbv for values less then 500 ppbv and by 50% for 500 ppbv or greater values, while for CO it is increased by 50 ppbv or a factor of two, whichever is larger. The impacts of these modifications are negligible compared to other systematic uncertainties modeled.

The potential impact of some remaining (typically small) systematic uncertainties has been quantified through analytic calculation based on simplified models of the MLS measurement system [*Read et al.*, 2007]. These calculations provide only an estimate of the 'gain uncertainty' (i.e., possible multiplicative error) introduced by the source in question, this approach does not quantify possible biases or additional scatters for these minor sources of uncertainty.

2.5.2. Results Figures 5 and 6 summarize the results of this quantification for UTLS O_3 and CO, respectively. These show the magnitudes of expected biases, additional scatters and possible scaling uncertainties the various errors may introduce into the data, and should be interpreted as 2-σ estimates of their likely magnitude.

The contribution of clouds to systematic uncertainty applies only to regions of thick cloud, and has been quantified by adding the effects of scattering from a representative cloud field to the simulated radiances. Retrievals based on these radiances, including the cloud radiance screening approach outlined in Section 2.3, have been compared to the unperturbed results. The bias and scatter shown are based on consideration of only the cloudy profiles (as defined by the known amount of cloud in the 'true' fields). In the case of UT/LS O_3, this study indicates a possible cloud-induced

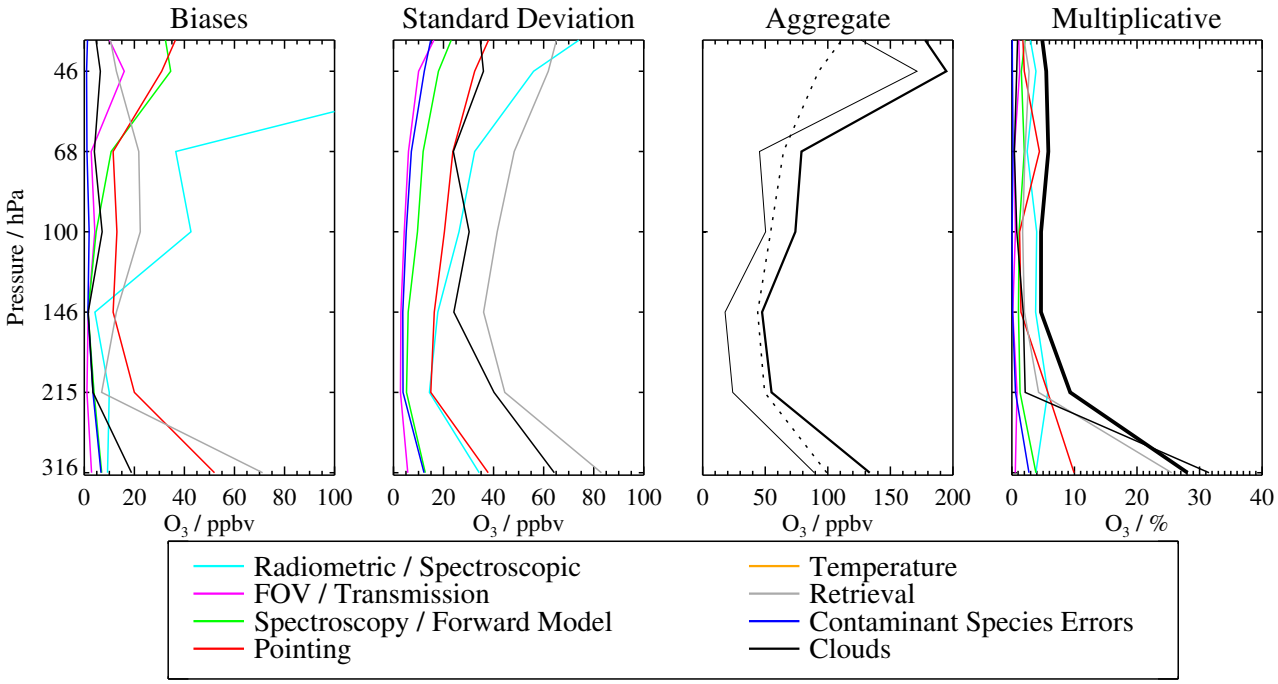

Figure 5. The estimated impact of various families of systematic errors on the MLS UTLS O_3 observations. The first two panels show the (first) possible biases, and (second) standard deviation of the additional scatter, introduced by the various families of errors, with each family denoted by a different colored line. Cyan lines denote errors in MLS radiometric and spectral calibration. Magenta lines show errors associated with the MLS field of view and antenna transmission efficiency. Red lines depict errors associated with MLS pointing uncertainty. The impact of possible errors in spectroscopic databases and forward model approximations are denoted by the green line, while those associated with retrieval formulation are shown in grey. The gold lines indicate possible errors resulting from errors in the MLS temperature product, while the blue lines show the impact of similar 'knock on' errors in other species. Finally, the typical impact of cloud contamination is denoted by the black line. (Third panel) the root sum squares (RSS) of all the possible biases (thin solid line), all the additional scatters (thin broken line), and the RSS sum of the two (thick solid line). (Fourth panel) the scaling uncertainties introduced by the various error sources (colors have the same meaning as for the first two panels). The thick black line shows the RSS scaling uncertainty.

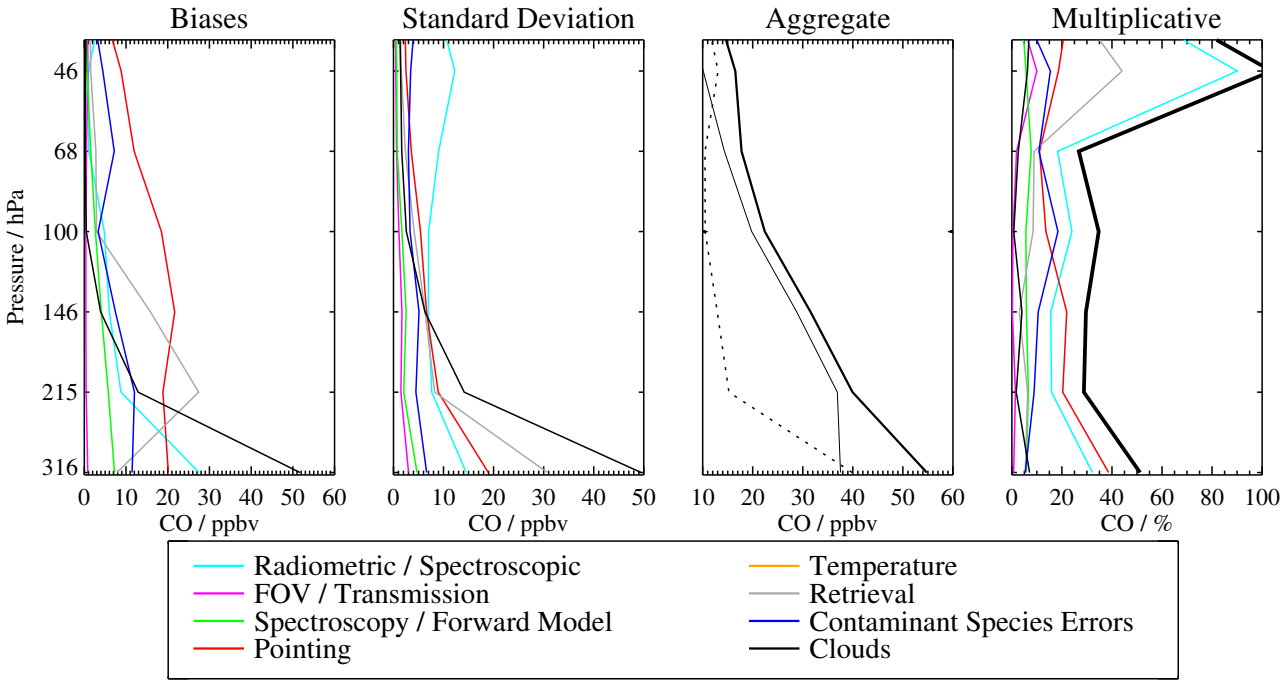

Figure 6. As Figure 5 except for UT/LS CO.

bias of $\sim \pm 5$ ppbv with an additional scatter of $\sim \pm 40$ ppbv at 215 hPa (less at smaller pressures). The corresponding impact on CO at 215 hPa is a bias of $\sim \pm 15$ ppbv with an additional scatter of $\sim$15 ppbv, with smaller impacts at lesser pressures. Both products show much larger cloud impacts (50 – 80 ppbv) for 316 hPa data.

The retrieval formulation uncertainty (grey lines) mainly reflect the difference between the retrieval of unperturbed simulated radiances and the 'true' model atmosphere. In the case of CO, a positive bias of $\sim$30$\pm$10 ppbv is seen at 215 hPa, thought mainly to be due to the handling of the spectrally flat 'extinction' terms in the forward model. This bias is likely to directly apply also to real MLS observations (unlike other uncertainty sources where the quoted value is a 2-σ estimate of the possible magnitude of any bias). The large bias and scatter this error source introduces into the 316 hPa O_3 data make it unlikely to be useful for scientific study.

Of the remaining uncertainty sources, those related to pointing issues (red), and MLS radiometric calibration (cyan) are the most significant, with contaminating species (blue) also being important for CO. The pointing uncertainties arise from a combination of the uncertainty in the width of the O_2 lines used to determine limb tangent pressure, and in uncertainty in the vertical offsets between the fields of view of the MLS 118 and 240-GHz receivers. The main component of the uncertainties associated with radiometric calibration originate from the spectral signature introduced in calibrated MLS radiances by departures from a linear response within the signal chains. In addition, standing waves within the MLS instrument contribute significantly to the systematic uncertainty in the 316 hPa CO data (O_3 data at this altitude are less affected).

Overall, this study indicates a potential bias of up to $\pm$25 ppbv for O_3 at 147 and 215 hPa, with an additional scatter of $\sim \pm$50 ppbv. For 100 hPa, the bias and scatter are $\sim \pm$50 ppbv each. Possible multiplicative errors in UT/LS O_3 are 10% at 215 hPa and 5% at smaller pressures. In the case of CO, there are potential biases of roughly $\pm$40, $\pm$30, and $\pm$20 ppbv at 215, 147 and 100 hPa, respectively $\sim \pm$10 ppbv. Possible scaling errors in the UT/LS CO product are around 30%. In scientific studies the accuracy quoted for each MLS data point should be the estimated bias plus the multiplicative error times the retrieved value. These findings are summarized, along with precision and resolution information in Tables 2 and 3. In some cases, the estimated accuracy quoted is based on comparisons with observations shown later in this paper, rather than on the expectations from this study.

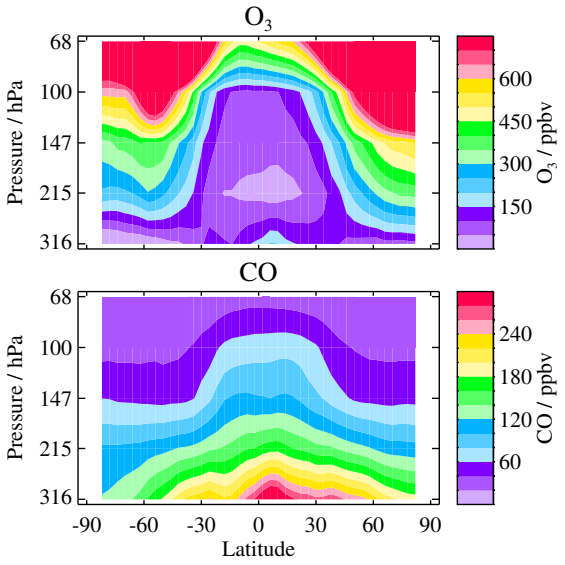

Figure 7. Zonal means from ∼80 days (distributed across seasons and years) of MLS v2.2 O_3 and CO data in the UT/LS region.

3. 'Zero order' validation of MLS UT/LS O_3/CO

3.1. Overview and comparisons with expectations

Figure 7 shows a zonal mean of ∼80 days of MLS v2.2 O_3 and CO data (distributed roughly evenly among years and seasons). The fields show the generally expected structure, with larger ozone abundances seen in or closer to the stratosphere. However, the 316 hPa O_3 values (not recommended for scientific use) show an unexpected peak in the tropics. The CO shows the expected morphology with low values in the stratosphere and generally larger values lower in the atmosphere. However, the absolute values appear too high compared to expectations at 215 and 316 hPa (as shown later, in situ observations indicate that, while abundances above 150 ppbv at these altitudes are possible, average values are more typically 50 – 100 ppbv).

Figure 8 compares these zonal means (and the standard deviations about them) to data from the MOZAIC commercial aircraft dataset [*Marenco et al.*, 1998; *Thouret et al.*, 1998; *Nédélec et al.*, 2003; *Nedelec et al.*, 2005]. The MOZAIC observations rarely extend to pressures smaller than 200 hPa. There is encouraging agreement between MLS O_3 and MOZAIC observations at 215 hPa. However, MLS CO data at these altitudes exhibit a high bias compared to MOZAIC. The CO data at 316 hPa also show a high bias and more latitudinal structure than is seen in MOZAIC data, while the 316 hPa MLS O_3 data show very little relationship to MOZAIC observations and generally very unexpected behavior. In all these comparisons, the strong emphasis of mid-latitude northern hemisphere observations in the MOZAIC dataset should be borne in mind.

MLS radiance observations in the upper troposphere in the 240 GHz region are dominated by emission from ozone, as shown in Figure 1. The poor quality of the MLS 316 hPa O_3 observations implies an inability of the MLS retrievals to correctly interpret the radiances measured at tangent pressures from ∼250 – 316 hPa that influence the 316 hPa retrievals. This in turn implies that the 316 hPa MLS upper tropospheric CO data is unlikely to be of sufficient quality for scientific use, despite having more reasonable morphology (though with a clear high bias). The same inference applies to the 316 hPa MLS HNO_3 observations [*Santee et al.*, 2007].

The combination of this inference with the results of the systematic error study in section 2.5 and the unusual form of the 316 hPa CO averaging kernel (see Figure 4) lead to the conclusion that version 2.2 O_3 and CO at 316 hPa are not suitable for scientific use.

Figure 9 shows histograms of MLS UT/LS O_3 and CO observations from ∼40 days of observations. The figure shows that the screening by 'Quality' and 'Convergence' described in section 2.2 discards some unrealistically small (often negative) values of O_3 and CO at 215 hPa in the tropics. These poor retrievals probably reflect poorly modeled cloud signatures in the MLS radiances. The ozone histograms clearly show the influence of stratospheric air at high and mid-latitudes. The data at 316 hPa (not recommended for scientific use) show a significant tail at higher values. While generally good agreement is seen between MLS and MOZAIC O_3 (with the exception of the tropics at 316 hPa), the CO histograms show the MLS data to be generally high compared to MOZAIC, with a generally larger dynamic range of values, particularly at 316 hPa. In the mid-latitudes around 215 hPa (where the bulk of the MOZAIC data were taken) a significant fraction of the CO observations indicate abundances larger than 150 ppbv. The MLS histogram in this region shows a somewhat similar tail, albeit with a clear bias towards higher values. The fact that a similar tail is not seen in MOZAIC data at 215 hPa in the tropics (a region where convective transport of polluted air is likely to be more frequent) may simply reflect the highly sparse nature of MOZAIC observations at these latitudes.

3.2. Comparisons with the GEOS-CHEM model

Figures 10 and 11 compare a week of the MLS v2.2 UT/LS O_3 and CO data to results from the GEOS-CHEM model [*Bey et al.*, 2001] run in a 'near-real-time' configuration (using climatological emissions and meteorology from

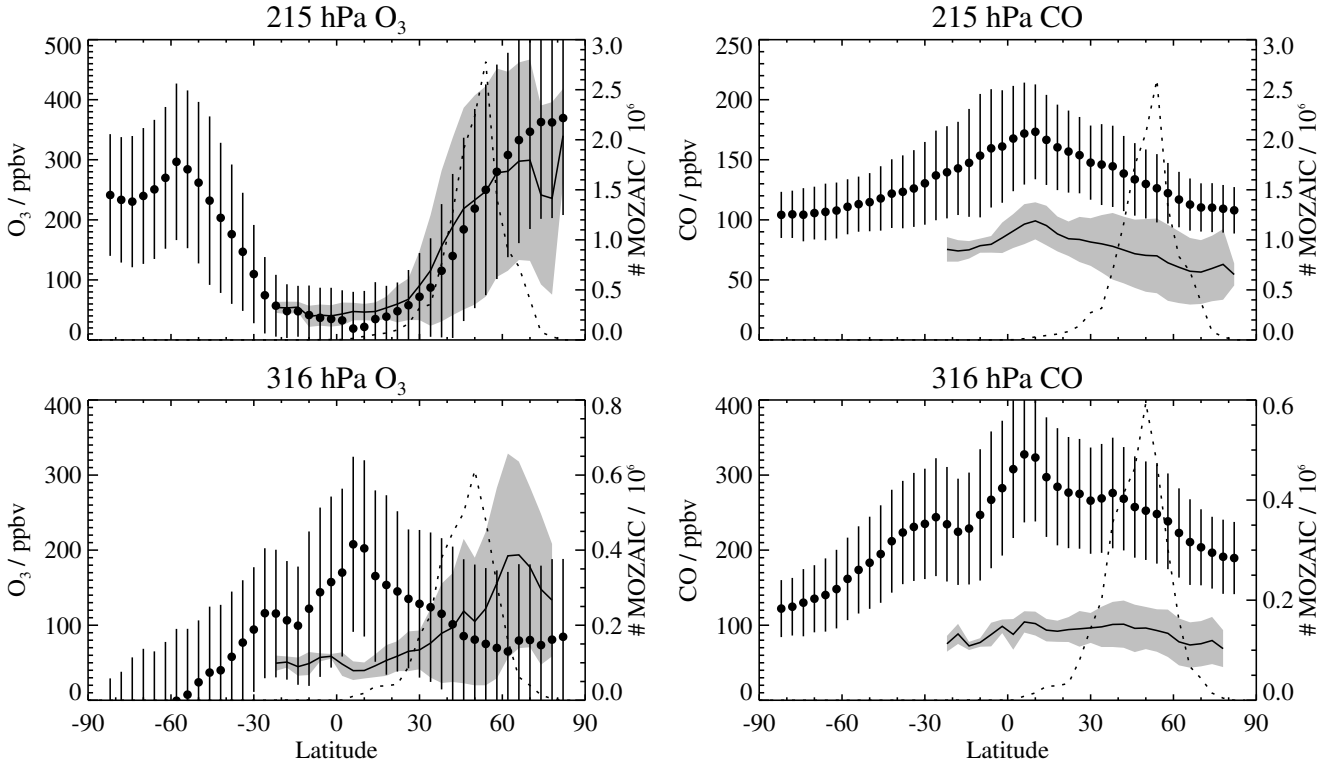

Figure 8. Zonal means (left hand axes) of MOZAIC (left panels) O_3 and (right panels) CO data for 2004 and 2005 (solid line plus grey shading indicating standard deviation), compared to the MLS zonal means shown in Figure 7 (points with vertical lines indicating standard deviation). The broken lines indicate the number of MOZAIC measurements that formed the averages (right hand axes). The '316 hPa' MOZAIC data is an average of all the measurements from 383 –261 hPa, while the '215 hPa' values are the 261 – 177 hPa average. Uncertainties on the MOZAIC observations are 2 ppbv precision, 2% accuracy for O_3, and 5 ppbv precision, 5% accuracy for CO.

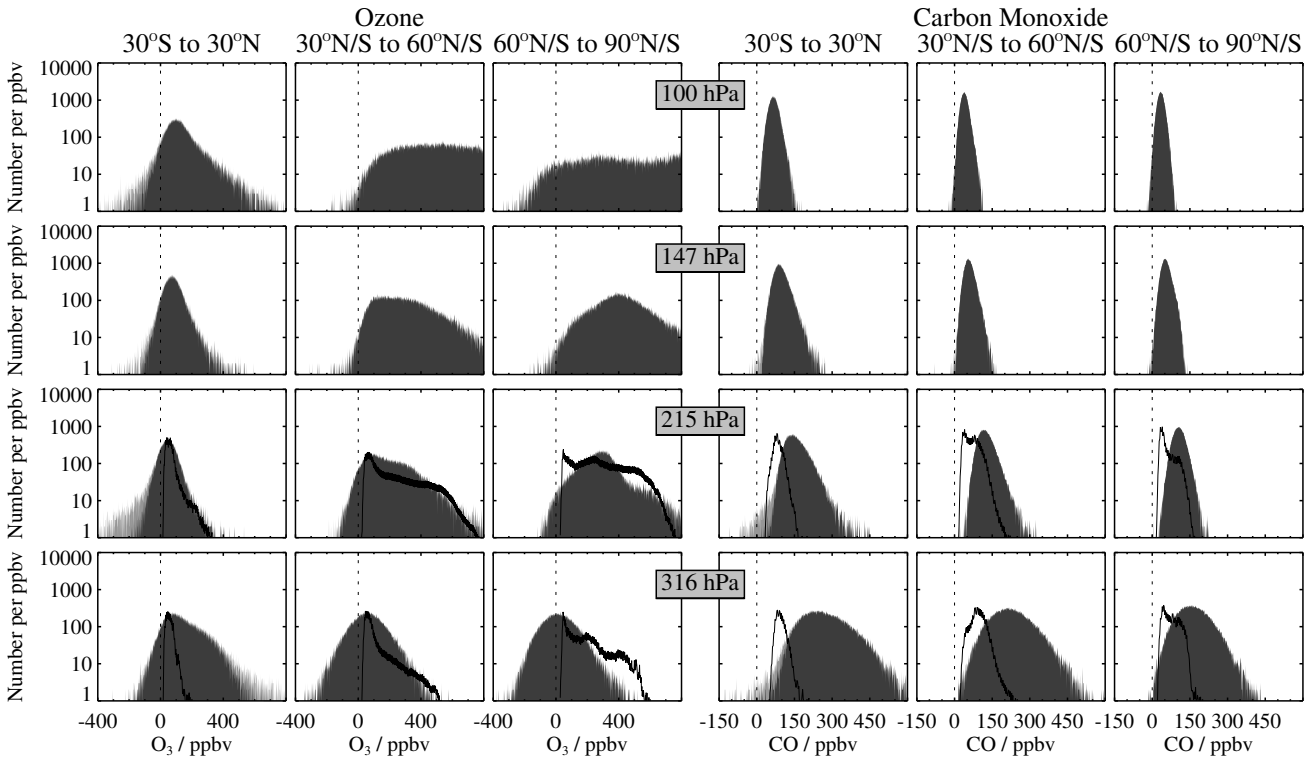

Figure 9. Histograms of all the MLS UT/LS O_3 and CO data for ∼80 days (distributed across seasons and years) in three latitude bins (tropics, mid-latitudes and polar regions). The light grey histogram shows all the MLS data for which the 'Status' field is an even number and the precision field is positive. The darker grey region shows those values meeting the 'Quality' and 'Convergence' screening described in section 2.2. The black lines show the comparable histogram for 2004 – 2005 MOZAIC data (renormalized vertically to fit the same scale).

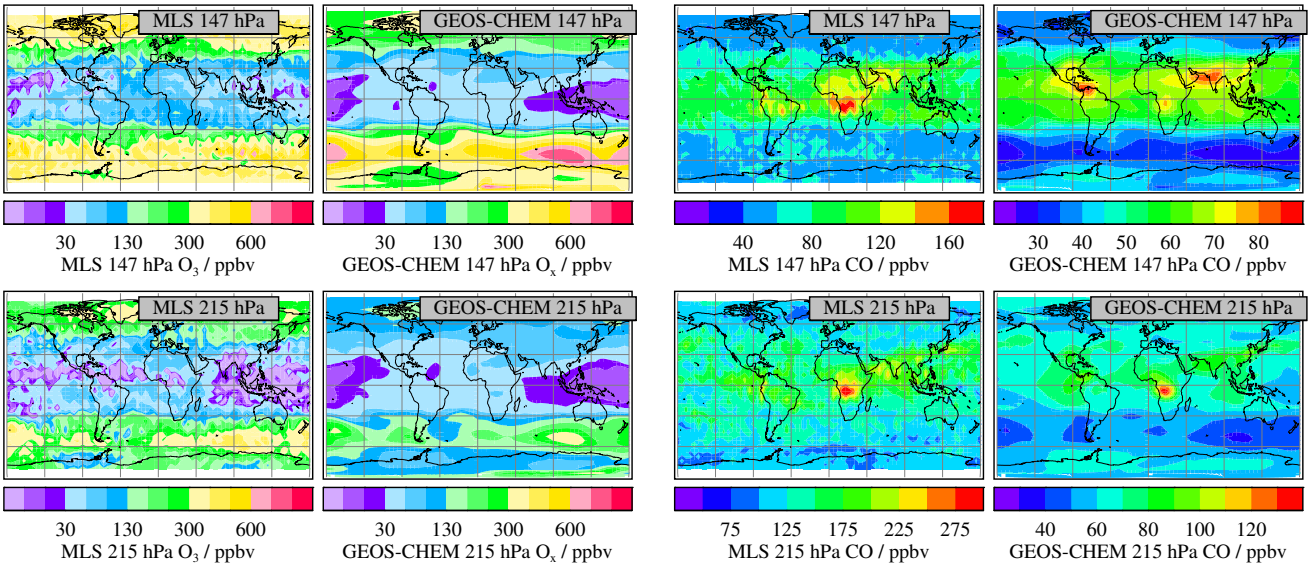

Figure 10. Comparison of MLS v2.2 and GEOS-CHEM (near real time run) 17–24 September 2005 average O_3 (left 2 columns) and CO (right two columns). Note the use of a common non-linear scale for the O_3 maps and varying scales for CO.

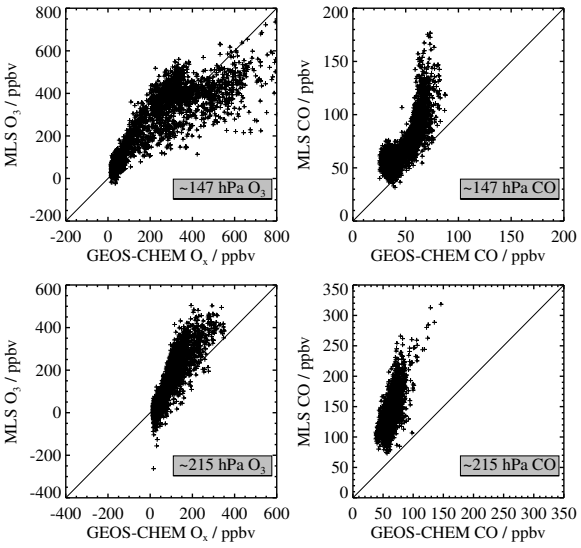

Figure 11. Detailed comparison of the MLS and GEOS-CHEM data shown in Figure 10.

the GEOS-4 dataset [*Bloom et al.*, 2005]). The MLS and model O_3 fields show reasonable agreement in morphology, with low values over the tropical western Pacific region compared to other longitudes. There is significant disagreement in O_3 abundances at higher latitudes. This probably reflects known limitations in the ability of GEOS-CHEM to model the stratosphere (being designed as a tropospheric chemistry model).

The MLS CO data are clearly higher than GEOS-CHEM at both altitudes shown. There is good agreement in morphology in the ∼215 hPa data. At 147 hPa the MLS data show strong enhancement in central Africa (as at 215 hPa) which is not reported by GEOS-CHEM. Away from this region the agreement in morphology between MLS and GEOS-CHEM at 147 hPa is reasonable. The differences at 147 hPa over central Africa may indicate that GEOS-CHEM underestimates the altitude to which convective transport can loft boundary-layer air. Similar differences in convective outflow altitudes have been seen in MLS v1.5 and GEOS-CHEM CO data throughout the Aura mission.

3.3. Validity of MLS CO morphology

While the v2.2 215 hPa MLS CO data show reasonable morphology, their high bias compared to MOZAIC and expectations indicates that the validity of this morphology bears further investigation. Study of the raw MLS radiance measurements is one way in which to gain confidence in the

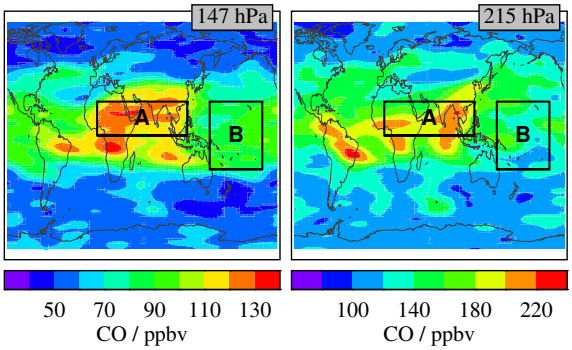

Figure 12. Maps of average v2.2 MLS CO at 147 hPa and 215 hPa from the period 17 – 19 September 2004.

validity of the data derived from them. As shown in Figure 1, the 240 GHz region observed by MLS is dominated by emission from ozone. In addition, a spectrally flat 'baseline' signature in the radiances comes from continuum emission by dry air and water vapor. The ∼1 K brightness temperature CO feature is small compared to these signatures, and is only discernible after these other contributions are characterized and removed.

Here, we repeat the analysis performed in *Filipiak et al.* [2005] (for a different time period) and extend it to consider MLS CO observations at 215 hPa. MLS v2.2 CO observations for 17 – 19 September 2004 are shown in Figure 12. Clear enhancements in CO at 147 and 215 hPa are seen over central Africa and southern Asia (region A), and also off the west coast of central America. Our analysis considers MLS radiance observations in this period from band 9 (the CO band), centered at 230.543 GHz in the lower sideband, and 248.777 GHz in the upper sideband. The CO spectral line is at 230.538 GHz (the ∼5 MHz offset compensates for orbital Doppler shift). The contributions from dry air and water vapor continua to the band 9 signal can be largely removed by subtracting the signal seen in a nearby window region (MLS band 33 channel 3 at ∼238.1 GHz/247.4 GHz). The remaining dependence of the band 9 signal on ozone can be estimated from the behavior of the MLS radiances in the 25°S – 25°N, 150°E – 150°W region (region B in Figure 12), where MLS reports lower abundances of CO with little significant morphology. In this region, the band 9 radiances show very good correlation with those in channel 25 of band 7, a channel sensitive to UT/LS O_3 but insensitive to CO abundances and only weakly sensitive to stratospheric ozone. Correlation coefficients are generally greater than 0.85, except for channels closer to the CO line center, which are more strongly affected by mesospheric CO signatures. Accordingly, a least-squares linear fit between each band 9 channel and band 7 channel 25 in this region can be used to deduce and subtract the band 9 ozone signature in other regions.

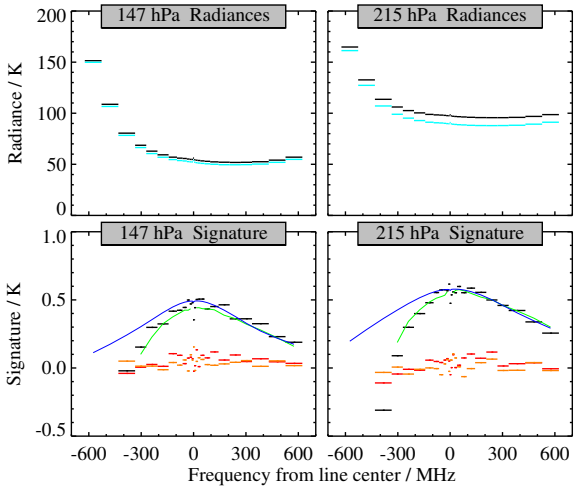

Figure 13. Top: Comparison of average MLS radiances measured in region A of the maps in Figure 12 (black) with those in region B (cyan). Bottom: Black lines show region A minus region B radiance differences when the signatures of continuum emission and upper tropospheric ozone have been removed as described in the text. The green lines show the same difference for the fit to the measured radiances achieved by the v2.2 software. The blue line shows the expected shape of the CO contribution at these altitudes based on a linearized forward model. The departure of the blue line from the black and green lines at the low frequency end is due to the non-linear impact of the ozone signature. The orange and red lines show the difference in measured radiances between the northern and southern (red) and eastern and western (orange) halves of region B. In these cases no signature of CO is seen in the differences.

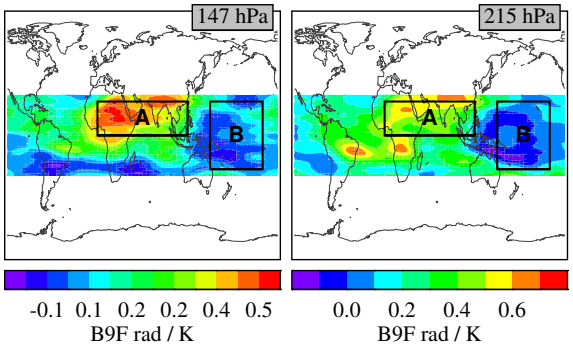

Figure 14. Maps of the CO radiance metric described in the text for the same period as shown in Figure 12.

Figure 13 compares average MLS radiance observations

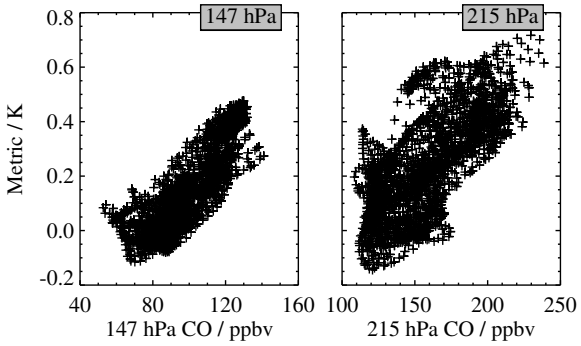

Figure 15. Comparison of the CO fields shown in Figure 12 with the radiance metric maps shown in Figure 14. The correlation coefficients are 0.83 and 0.71 respectively for 147 and 215 hPa.

in regions A and B and shows that the difference between these, once the continuum and ozone signatures are removed, has a clear CO signature. A simple metric of CO abundance can be defined as the average radiance in this signature seen in channels 1 – 7 and 16 – 21 (channels 8 – 15 are affected by strong emission from mesospheric CO, while channels 22 – 25 are more strongly affected by ozone, as described above). Figure 14 shows a map of this metric at 147 and 215 hPa for comparison with the maps in Figure 12, while Figure 15 compares the two fields directly. The agreement is generally good, with the morphology in this metric comparing well with that seen in the v2.2 MLS CO data. The scatter in the relationship between MLS CO and the metric reflects differences in the impact radiance noise has on these two quantities. The fitting of the O_3 signature based on tropical radiances (region B) is not generally applicable to other latitudes due to differences introduced from changes in stratospheric ozone and the contributions of the moist and dry-air continua. This metric is therefore only applicable to tropical areas.

3.4. Comparison with MLS v1.5 data

This paper describes MLS UT/LS O_3 and CO data produced by version 2.2 of the MLS data processing algorithms. The previous version of MLS data, version 1.5, has been produced for the majority of days from August 2004 to the end of February 2007, and has been used in a large number of scientific studies. Version 1.5 CO data have formed the basis of several scientific papers, including the discovery of a 'tape recorder' signal [*Schoeberl et al.*, 2006a], a study of transport paths into the stratosphere [*Fu et al.*, 2005], quantification of the influence of convection on upper tropospheric composition [*Folkins et al.*, 2006], and the trapping of polluted air in the upper troposphere [*Li et al.*, 2005].

The main difference between the v1.5 and v2.2 UT/LS O_3 and CO data arises from the manner in which the retrieval models spectrally 'flat' features in the 240 GHz MLS radiances. (Spectrally flat in this context means that they exhibit no significant variation across the few GHz-wide signatures of O_3 and CO.) MLS measurements are based on observations of spectral contrast; the 'baseline' on which the observed spectral lines sit is generally of less concern. The v1.5 algorithms fitted a spectrally flat 'baseline' to each observed MLS radiance, an approach that works well for situations (such as the MLS 640 GHz observations) where these spectrally flat signatures arise from instrumental sources.

By contrast, as discussed in *Livesey et al.* [2006], signatures that are locally spectrally flat in the atmosphere do not necessarily have a spectrally flat impact on the MLS radiances because of variations in atmospheric optical depth with frequency (some channels see deeper in the atmosphere than others). Such signatures can arise as a result of uncertainty in water vapor abundance or continuum models, or from the scattering of radiances by thick clouds (though, as discussed, the latter can have some spectral impact in extreme cases). For the MLS 240 GHZ UT/LS observations these effects are more significant than instrumental effects. Instead of retrieving 'baseline', a better approach to compensating for these effects is to retrieve vertical profiles of spectrally flat atmospheric extinction. This is the approach taken in version 2.2.

Figures 16 and 17 compare 1 May 2006 v1.5 and v2.2 data for UT/LS O_3 and CO, respectively. The O_3 data show generally good agreement between the two versions. The CO, by contrast, is markedly improved in v2.2 over the earlier v1.5 product. Radiance signatures of thick clouds led to very high anomalous values of CO in v1.5, many of which were not identified as suspect by the retrieval algorithms. In v2.2 there are far fewer anomalous values and very few clearly unrealistic values not identified as bad by the retrieval. While v1.5 reported very high values of CO in cases of cloud contamination, v2.2 generally reports low or negative values (somewhat similar behavior is seen in the O_3 product). In general, the v2.2 CO data show less difference between the tropics and mid-latitudes at 147 hPa than is seen in v1.5. This change has brought the data into better agreement with models such as GEOS-CHEM. The underlying high bias at 215 hPa CO was apparent also in the earlier v1.5 data, where it was enhanced by the anomalously large values associated with unreported cloud contamination.

4. Comparisons with other observations

The validation of Aura observations has been the focus, or partial focus, of significant aircraft campaigns since the

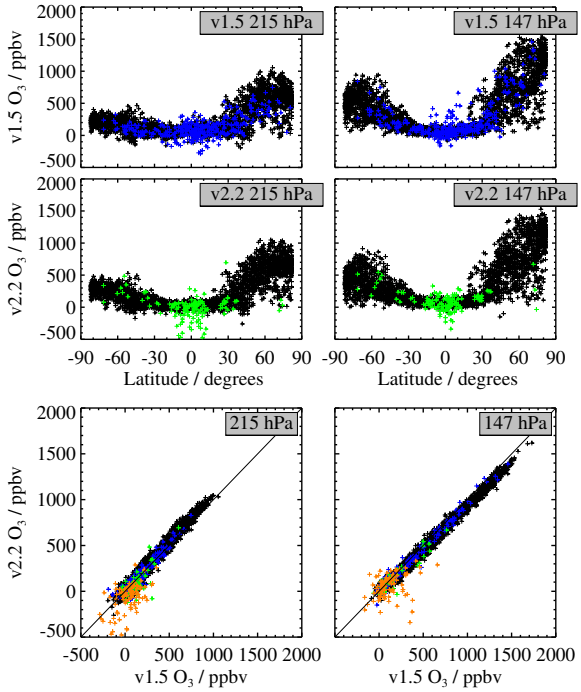

Figure 16. Comparison of MLS v1.5 and v2.2 O_3 data for 1 May 2006. The left hand plots show data at 215 hPa, with 147 hPa data shown on the right. The top row shows MLS v1.5 O_3 data as a function of latitude. Black points meet all the screening criteria given for v1.5 O_3 data (*Livesey et al.* [2005]). Blue points are those v1.5 O_3 points where quality>0.1 or status was a non-zero even number, indicating cloud contamination (points with negative precision or odd values of status are completely neglected in all these plots). The second row of plots shows the equivalent for v2.2 O_3, with the green points being those where quality>1.2 or convergence<1.8 (see Section 2.2, the status=0 criteria is not required for v2.2 data). The final row of plots scatters the v2.2 data (y-axis) against v1.5 (x-axis) with black points indicating data that was 'good' in both datasets, blue and green indicating data marked bad by v1.5 and v2.2, respectively, and orange symbols used for points identified as 'bad' in both versions.

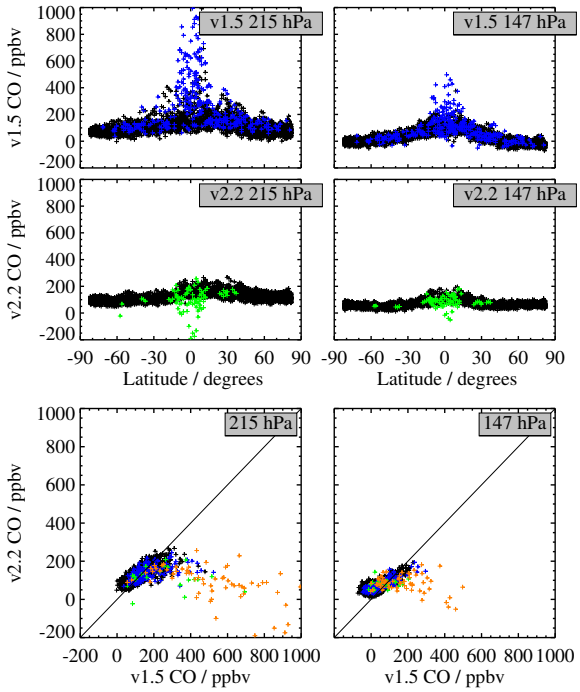

Figure 17. As Figure 16 but for CO. Here the v1.5 screening criterion was quality>0.05, the v2.2 screening criteria are the same as for O_3.

Aura launch. The Polar Aura Validation Experiment (PAVE) during January/February 2005 consisted of multiple flights of the NASA DC-8 to high latitudes over North America. Measurements from this campaign include vertical profiles of ozone above and below the aircraft from Light Detection and Ranging (LIDAR) instruments and in-situ observations of both O_3 and CO. The April/May 2006 Intercontinental Chemical Transport Experiment (INTEX-B) campaign made similar observations (among many others) from the same aircraft in the Northern Pacific (generally at lower altitudes). The Houston deployments of the Aura Validation Experiment (AVE) in October/November 2004 and January/February 2005 and the Costa-Rica AVE (CR-AVE) deployment in March 2006 provided in-situ measurements of several MLS species, including O_3 and CO from the NASA high-altitude WB-57 aircraft in the tropical upper troposphere and lower stratosphere.

Comparisons with data from other satellite sensors are also possible in the UT/LS region. *Froidevaux et al.* [2007] compares MLS v2.2 O_3 data with observations from the Stratospheric Aerosol and Gas Experiment (SAGE) instruments. Similarly, *Pumphrey et al.* [2007] compare MLS v2.2 CO data with observations from the Canadian ACE satellite. In addition to these, comparisons of MLS v2.2

UT/LS O_3 and CO with observations from the Aura Tropospheric Emission Spectrometer (TES, *Beer* [2006]) and the Aqua Atmospheric Infrared Spectrometer (AIRS, *Aumann et al.* [2003]) are possible, as are comparisons with CO observations from the Measurement of Pollution in the Troposphere (MOPITT, *Drummond and Mand* [1996]) instrument on Terra. Such comparisons are complicated by the broad nature of the vertical averaging kernels for these observations, including significant sensitivity in the lower troposphere not measured by MLS. As such, these studies are beyond the scope of this paper, though preliminary comparisons are underway.

4.1. Comparisons with airborne LIDAR O_3 data

During the PAVE and INTEX-B campaigns, the DC-8 payload included two LIDAR instruments measuring ozone. The Differential Absorption Lidar (DIAL) instrument [*Browell et al.*, 1990, 1998] observes ozone above and below the aircraft, while the Airborne Raman Ozone, Temperature and Aerosol LIDAR (AROTAL) looks only upwards [*McGee et al.*, 1993, and references therein]. The AROTAL and DIAL observations during PAVE were focused on validation of stratospheric ozone, and are discussed in *Froidevaux et al.* [2007]. This paper considers the DIAL measurements of O_3 during INTEX-B.

Figure 18 compares MLS O_3 observations with those from DIAL [*Browell et al.*, 1998, 1990]. The aircraft data shown here also include in situ observations from the FASTOZ instrument [*Pearson and Steadman*, 1980; *Eastman and Steadman*, 1977] and interpolation to fill in data for the regions immediately above and below the aircraft, where the LIDAR provides no information. For those portions of the INTEX-B flights that were along the MLS track, the DC-8 was generally flying around 200 hPa. Accordingly, the FASTOZ data contributes significantly to combined LIDAR / in-situ dataset used in these comparisons.

In making these comparisons, it is important to bear in mind that MLS data do not represent 'layer means', rather they define piecewise-linear profiles in pressure that best match the observed radiances [*Read et al.*, 2006]. This piecewise linear representation also applies in the along-track direction. Accordingly, the most appropriate manner in which to compare MLS data to high resolution LIDAR measurements is to perform a least-squares fit of the LIDAR data onto the MLS 'grid points' [*Livesey et al.*, 2006]. The second panel of Figure 18 shows the results of this fit. The relative dearth of LIDAR observations at pressures smaller than 147 hPa makes the fits at 147 and, especially, at 100 hPa under-constrained, and has clearly led to extrapolated values at 100 hPa, resulting in artificially large differences between DIAL and MLS at 100 hPa (bottom panel).

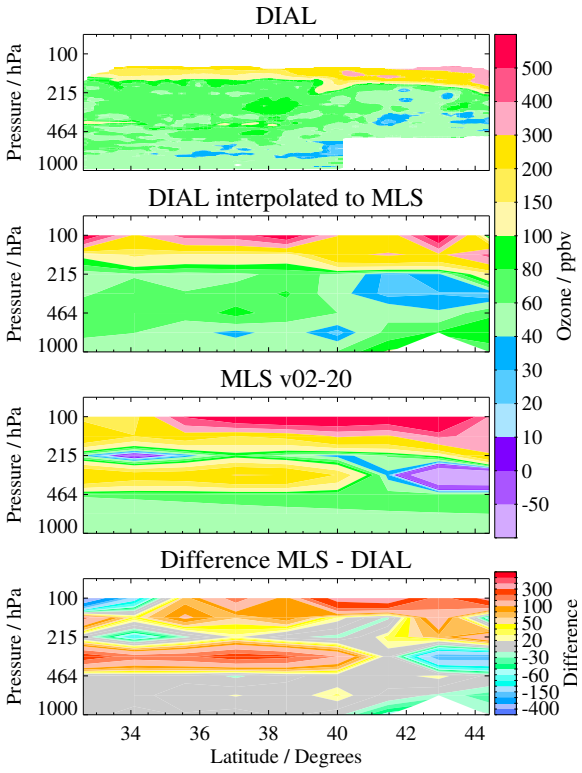

Figure 18. Comparison of DIAL LIDAR and MLS UT/LS O_3 observations on 25 April, 2006. The top panel shows the DIAL LIDAR observations (mapped onto pressure surfaces using the MLS geopotential height data [*Schwartz et al.*, 2007]). The second panel shows a least-squares fit of this field to the MLS retrieval grid (6 surfaces per decade change in pressure, with profiles every ∼165 km great circle distance horizontally). The third panel shows the coincident MLS data and the final panel shows the difference between the fitted DIAL and MLS.

To help identify such cases, the least-squares fit is weighted, with all LIDAR data given an equal 1 ppbv weighting. An a priori field of 0 ± 1000 ppbv is included in the fit to stabilize the matrix inversion. If this were not done, output grid points not influenced by any LIDAR observations would fail to invert. The weighted least squares fit provides an uncertainty estimate on each derived grid point. As each grid point typically represents a fit to several individual LIDAR data points, the estimated precision on the fitted grid points is typically $\sim$0.08 ppbv. Precisions of worse than 0.1 ppbv indicate fits based on little LIDAR information, which should therefore be ignored. (Note that the absolute values of the weights chosen here are immaterial; it is the ratio of the LIDAR and a priori weights that is critical.)

Figure 19 shows a comparison of MLS observations with the results of this fit for all the INTEX-B DIAL coincidences. This shows good agreement, notably for high values at 215 hPa, mainly corresponding to stratospheric observations from the 7 May, 2006 flight. The 316 hPa data (not recommended for scientific use) show less agreement, and the agreement of the few useful points at 147 hPa is encouraging but not definitive. The large scatter seen in these comparisons compared to the MLS error bars may reflect atmospheric variability seen by MLS but not captured by the aircraft. This includes variations across the MLS line of sight (i.e., perpendicular to the DC-8 flight track) and unsampled variability at the altitudes where the DC-8 data represents in interpolation between the DIAL lidar and the FASTOZ in-situ measurements.

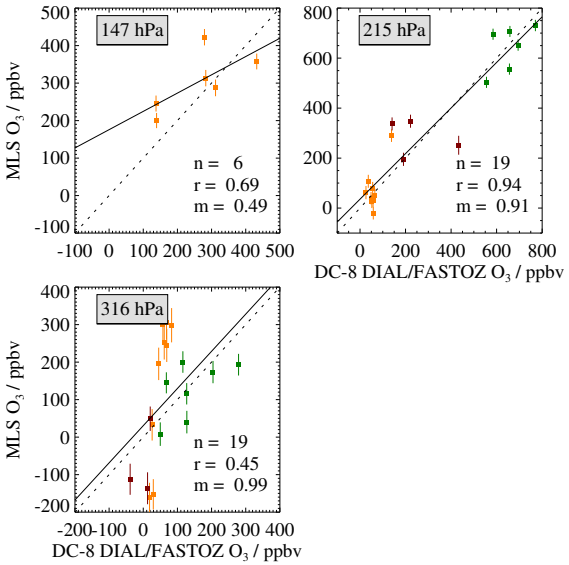

Figure 19. Summary of all the MLS/DIAL LIDAR ozone comparisons with separate panels for each MLS retrieval surface. Each data point compares one MLS retrieved O_3 value (y-axis) with an estimate from a least-squares fit to INTEX-B LIDAR data. The vertical error bars denote the MLS precision estimates. Different colors are used to denote different DC-8 flights, see Figure 26 for details. The number of points (n), correlation coefficient (r) and linear fit gradient (m) are quoted for each panel.

4.2. Comparisons with in-situ aircraft data

4.2.1. WB-57 Ozone comparisons The WB-57 flights during the various AVE campaigns provided several opportunities for comparisons of MLS UT/LS O_3 and CO with in situ observations. Figure 20 shows a comparison of O_3 measured by MLS and the NOAA ozone instrument on 22 January 2006. While the least-squares fit approach used to map LIDAR observations to the MLS grid is applicable to in situ observations, such fits are very unstable, because of the sparse nature of the aircraft observations. Instead, for all in situ comparisons shown here, we simply compare the MLS data to the average (in mixing ratio space) of all the in situ data points that fall within a 6/decade vertical, 1.5°great circle angle "box" centered on the MLS point, as shown by the colored boxes in the upper panel of Figure 20. Points where the aircraft departed from the MLS measurement track by more than 100 km or 24 hours were discarded.

The extent to which the multiple aircraft observations within a "box" are representative of the average mixing ratio in that box (the quantity most comparable to individual MLS observations) is hard to quantify, as the amount of at-

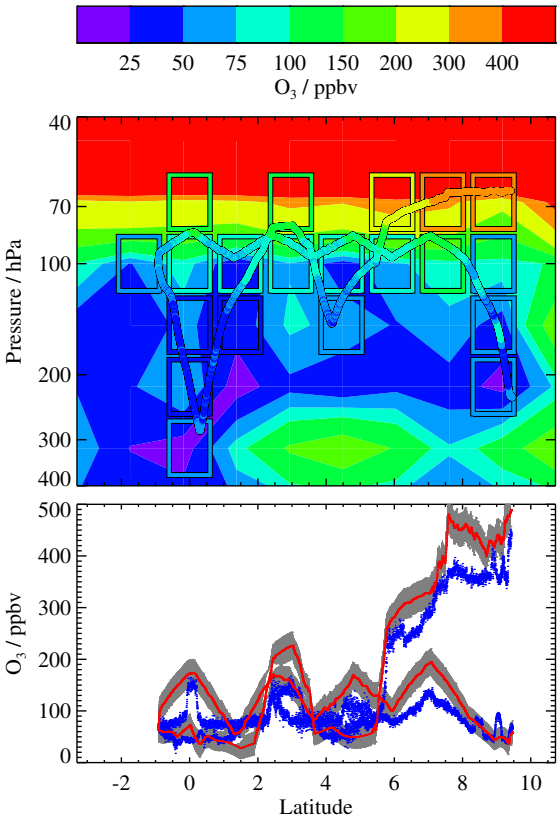

Figure 20. Comparison of MLS with WB-57 O_3 observations from the NOAA instrument on 22 January 2006. Top: The colored contour field shows MLS retrieved O_3 as a function of pressure and latitude, while the overlaid black bordered dots (merging to form a continuous trace) show WB-57 observations on the same color scale. The colored boxes show the "averaged" WB-57 observations to be compared to the MLS data, as described in the text. Bottom: The WB-57 data versus latitude (blue) compared to the MLS fields interpolated (in pressure and latitude) to the aircraft location (red with 1-σ precision shown in grey).

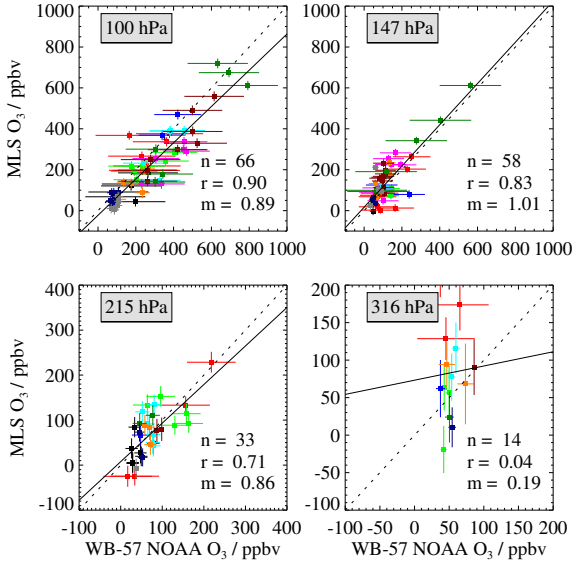

Figure 21. Summary of all the MLS/WB-57 ozone comparisons, with separate panels for each MLS retrieval surface. The points are individual MLS O_3 values (y-axis) compared to average nearby in situ observations (x-axis). The vertical error bar shows the estimated precision on the MLS O_3 while the horizontal bar is related to the variability seen in the WB-57 observations as described in the text. Different colors denote different flights, as shown on Figure 24. The solid line represents a least-squares linear fit to the data and the broken line shows the one-to-one correlation line. The number of points (n), correlation coefficient (r) and linear fit gradient (m) are quoted for each panel.

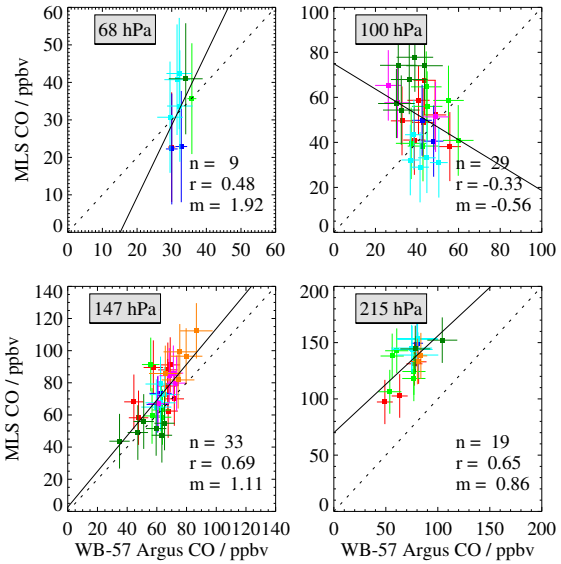

Figure 22. As Figure 21 except comparing MLS CO with WB-57 Argus observations. The estimated absolute accuracy for Argus data is 2%, traceable to CMDL standards.

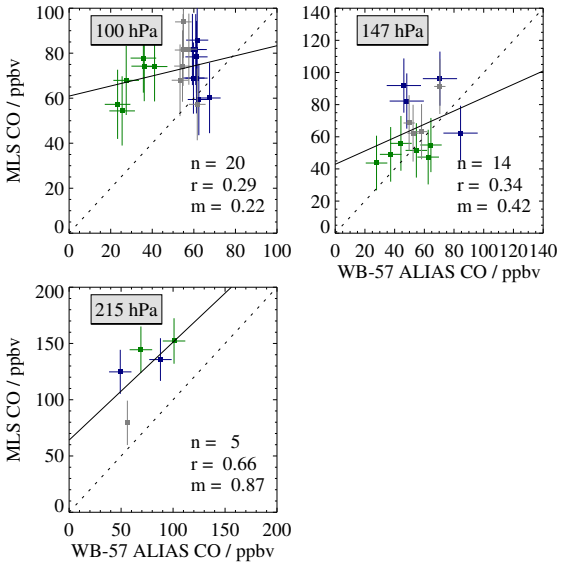

Figure 23. As Figure 21 comparing MLS CO to measurements from WB-57 ALIAS measurements during the Costa-Rica AVE campaign.

mospheric variability throughout the box is unknown. Simply taking the standard deviation of the in situ data within the box would lend undue weight to those cases where the aircraft sampled only a small fraction of the box. Instead, the uncertainty reported (shown by the horizontal error bars in Figure 21) is the largest standard deviation (i.e., variability) seen by the aircraft within any of the boxes at a given pressure level for a given flight. Figure 21 summarizes all the MLS/WB-57 O_3 comparisons by this method and shows very encouraging levels of agreement. The least-squares linear fit to the comparisons (excluding 316 hPa data) shown by the solid line in Figure 21 has a gradient of 0.79 with a correlation coefficient of 0.86. The normalized χ^2 statistic for the 1-1 fit is 0.12, implying statistically significant biases in this comparison. However, such a low value indicates that we have probably overestimated the uncertainties in this study.

4.2.2. WB-57 and DC-8 CO comparisons The WB-57 payload also included the Argus and 'Aircraft Laser Infrared Absorption Spectrometer' (ALIAS) [*Webster et al.*, 1994] instruments, which provided CO observations. The PAVE and INTEX-B missions included CO observations from the Differential Absorption CO Measurements (DACOM) instrument. Figures 22, 23 and 25 summarize these in the same manner as Figure 21.

Unlike for O_3, the observed variability in UT/LS CO is small compared to the ∼20 ppbv precision on individual MLS data points. This makes it hard to draw definitive conclusions from the in situ comparisons. Many of the Argus

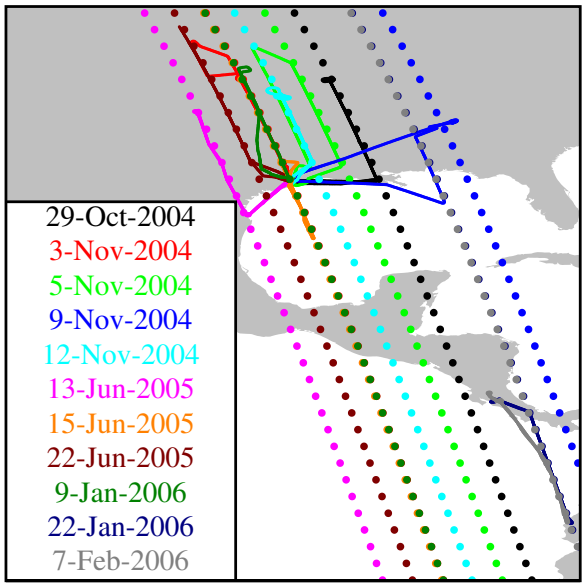

Figure 24. Map showing all the WB-57/MLS coincidences during the AVE missions. Colors are as used in Figures 21, 22 and 23.

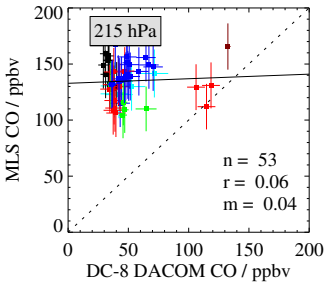

Figure 25. As for Figure 21 except comparing MLS CO to observations from DACOM on the NASA DC-8. Only comparisons for the 215 hPa MLS surface are shown as this level is close to the DC-8 flight ceiling. The large bias seen here is consistent with the Argus and ALIAS comparisons at 215 hPa indicating a factor of 2 high bias. The colors used for the different flight days are described in Figure 26, except for Magenta which is for a 22 March INTEX-B transfer flight.

and ALIAS comparisons at 147 hPa and smaller pressures lie within 1-σ of the one-to-one line, with the majority within 2-σ. The largest differences (50 – 100 ppbv) are seen in the 215 hPa comparisons. At these altitudes, MLS also consistently reports values ∼50–100 ppbv higher than are seen by DACOM. All the comparisons at 215 hPa (notably those with DACOM) show results consistent with the comparisons shown earlier (e.g., with MOZAIC and models) indicating a factor of ∼2 high bias in the MLS observations. There is a notable lack of correlation between MLS and DACOM for the 31 January 2005 PAVE and 1 May 2006 INTEX-B DC-8 flights. We note, however, that neither flight was targeting MLS validation and the coincidences are poorer than for other comparisons.

Drawing more quantitative conclusions is challenging. In all cases, χ^2 statistics imply no significant departs from a 1-1 relationship for the MLS data at 147 hPa or lesser pressures. Highly significant departures from a 1-1 relationship are indicated for all the 215 hPa MLS v2.2 CO comparisons, but no significant departures are seen from the individual linear fits. These comparisons highlight the difficulties associated with comparing measurements with disparate precisions and vastly different sampling volumes using only a few data points.

5. Summary, conclusions and future plans

Version 2.2 of the MLS data processing algorithms produce O_3 and CO profiles that are scientifically useful in the upper troposphere / lower stratosphere at pressures of 215 hPa and smaller. However, in the case of the 215 hPa

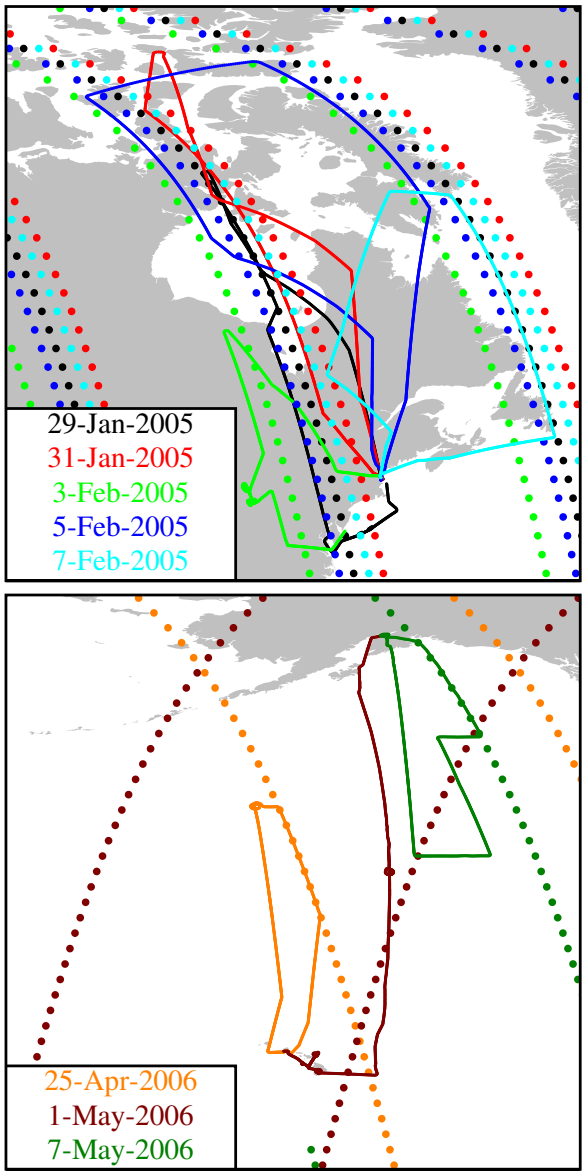

Figure 26. Maps showing the DC-8/MLS coincidences during the PAVE (top) and INTEX-B (lower) missions. Colors are as used in Figures 19 and 25. The 22 March 2006 flight in the INTEX-B campaign which was a transfer from Houston to NASA-AMES is not shown.

CO product, the observed factor of ∼2 high bias compared to other observations needs to be borne in mind, and studies generally limited to considerations of morphology only. The vertical resolution for the O_3 product is ∼2.5 km, with ∼4 km for CO. The cross-track horizontal resolution for both products is ∼6 km and along-track is ∼200 km for O_3, and ∼350 km for CO. Individual profiles (spaced ∼165 km along the MLS track) have a typical precision in the UT/LS of 20 – 40 ppbv for O_3 and 15 – 40 ppbv for CO. When using MLS data in scientific studies, care must be taken to screen the data according to the rules given in Section 2.2 [*Livesey et al.*, 2007].

The expected resolution, precision and accuracy of the MLS v2.2 O_3 and CO data are summarized in Tables 2 and 3, respectively. The accuracy to be used in scientific studies is given by the bias uncertainty plus the value times the scaling uncertainty. The overall error bar to use for a point is this accuracy plus the precision as scaled according to the number of points that go into a given average.

Comparisons with expectations, models and other observations generally corroborate the scientific usefulness of these products. The O_3 data show good agreement with expectations and observations, broadly in line with the systematic error study above. The most notable anomaly is the significant (factor of ∼2) high bias in the MLS CO product at 215 hPa. This is inconsistent with the systematic error budget described in Section 2.5 that estimated a possible ±40 ppbv, ±30% bias. This clearly indicates limitations in error quantification. This disconnect may lie in the assumptions of linearity used in the quantification of some of the smaller error sources. The MLS radiance signatures in the upper troposphere generally exhibit non-linear dependence on the atmospheric state. This issue will be the focus of future study.

The comparisons described here show no evidence for a bias in the MLS v2.2 O_3 at 215 hPa larger than ∼15%. However, we note that comparisons with SAGE [*Froidevaux et al.*, 2007] and radiosondes [*Jiang et al.*, 2007] indicate a ∼20% high bias in MLS v2.2 O_3 at 215 hPa at mid and high latitudes, although comparisons with ground based LIDAR [*Jiang et al.*, 2007] show good agreement in this region. Given these disagreements we ascribe an accuracy of ±20 ppbv and ±20% to the v2.2 MLS 215 hPa O_3.

The bias in the v2.2 MLS upper tropospheric CO data compared to other observations clearly needs further investigation. Further studies are needed to ascertain the nature of this bias (absolute offset, scaling error etc.). More detailed comparisons with MOZAIC and other satellite observations will yield more insight here. Understanding the origin of the CO bias and hopefully correcting it is an important goal for future versions of the MLS data processing algorithms. Planned research will also seek to extend the useful range of these data lower in the troposphere, and further reduce the sensitivity to contamination by thick clouds.

Acknowledgments. The research described in this paper was carried out by the Jet Propulsion Laboratory, California Institute of Technology, under a contract with the National Aeronautics and Space Administration. We are very grateful to the MLS instrument and data operations and development team for their support through all the phases of the MLS project, in particular D. Flower, G. Lau, J. Holden, R. Lay, M. Loo, D. Miller, B. Mills, S. Neely, G. Melgar, A. Hanzel, M. Echeverri, A. Mousessian and C. Vuu. We greatly appreciate the efforts of B. Bojkov and the Aura Validation Data Center (AVDC) team, whose work facilitated the MLS validation activities. Many thanks are due to the Aura Project for their support throughout the years (before and after Aura launch), in particular M. Schoeberl, A. Douglass (also as co-chair of the Aura validation working group), E. Hilsenrath, and J. Joiner. We also acknowledge the support from NASA Headquarters, P. DeCola for MLS and Aura, and M. Kurylo, J. Gleason, B. Doddridge, and H. Maring, especially in relation to the Aura validation activities and campaign planning efforts. The aircraft campaigns themselves involved tireless hours from various coordinators, including D. Fahey, E. Jensen, P. Newman, M. Schoeberl, H. Singh, D. Jacob as well as K. Thompson, and others involved with campaign flight management and support. The authors thank D. Edwards and L. Emmons for useful discussions. The authors acknowledge, for their strong support, the European Commission, Airbus and the Airlines (Lufthansa, Austrian and Air France) who, since 1994, have maintained and carried, free of charge the MOZAIC equipment. MOZAIC is supported by the "Institute National des Sciences de l'Univers – Centre National de la Recherce Scientifique", France, "Météo France" and "Forshungszentrum Jülich", Germany.

References

Aumann, H. H., et al. (2003), AIRS/AMSUB/HSB on the Aqua mission: Design, science objectives, data products, and processing systems, *IEEE Trans. Geosci. Remote Sens.*, *41*(2).

Beer, R. (2006), TES on the Aura mission: Scientific objectives, measurements, and analysis overview, *IEEE Trans. Geosci. Remote Sens.*, *44*(5), 1102–1105.

Bey, I., et al. (2001), Global modeling of tropospheric chemistry with assimilated meteorology: Model description and evaluation, *J. Geophys. Res.*, *106*, 23,073–23,096.

Bloom, S. C., et al. (2005), The Goddard Earth Observing Data Assimilation System, GEOS DAS Version 4.0.3: Documentation and validation, *Tech. Rep. NASA Tech. Rep. 104606 V26*, NASA.

Table 2. Summary of MLS UT/LS O_3 product

Pressure	Resolution[a] / km	Precision[b] / ppbv	Bias uncertainty / ppbv	Scaling uncertainty	Comments
46 hPa and less	—	—	—	—	See Froidevaux et al. (this issue)
68 hPa	3 × 200	±50	±50	±5%	
100 hPa	3 × 200	±40	±50	±5%	
147 hPa	3 × 200	±40	±20	±5%	
215 hPa	3 × 200	±40	±20	±20%[c]	
316 hPa	—	—	—	—	Unsuitable for scientific use
1000 – 464 hPa	—	—	—	—	Not retrieved

[a]Vertical, along-track. Cross-track resolution is ∼6 km

[b]Precision on individual profiles

[c]Indicated by comparisons with other observations described in this and other papers, rather than from the predictions used for other levels, see text for discussion

Table 3. Summary of MLS UT/LS CO product

Pressure	Resolution[a] / km	Precision[b] / ppbv	Bias uncertainty / ppbv	Scaling uncertainty	Comments
46 hPa and less	—	—	—	—	See Pumphrey et al. (this issue)
68 hPa	4 × 400	±10	±10	±30%	
100 hPa	4 × 500	±20	±20	±30%	
147 hPa	4 × 500	±20	±30	±30%	
215 hPa	5 × 600	±20	±40	∼ +100% [c]	See text
316 hPa	—	—	—	—	Unsuitable for scientific use
1000 – 464 hPa	—	—	—	—	Not retrieved

[a]Vertical, along-track. Cross-track resolution is ∼6 km

[b]Precision on individual profiles

[c]Indicated by comparisons with other observations, rather than from the predictions used for other levels, see text for discussion

Browell, E. V., S. Ismail, and W. B. Grant (1998), Differential absorption lidar (DIAL) measurements from air and space, *Appl. Phys. B*, *67*, 399–410.

Browell, E. V., et al. (1990), Airborne lidar observations in the wintertime Arctic stratosphere: Ozone, *Geophys. Res. Lett.*, *17*, 325–328.

Drummond, K. R., and G. S. Mand (1996), The Measurements of Pollution in the Troposphere (MOPITT) instrument: Overall performance and calibration requirements, *J. Atmos. Ocean. Technology*, *13*, 314–320.

Eastman, J. A., and D. H. Steadman (1977), A fast response sensor for ozone eddy-correlation flux measurements, *Atmos. Environ.*, *11*, 1209–1211.

Filipiak, M. J., et al. (2005), Carbon monoxide measured by the EOS Microwave Limb Sounder on Aura: First results, *Geophys. Res. Lett.*, *32*(14), L14,825, doi:1029/2005GL022765.

Folkins, I., P. Bernath, C. Boone, G. Lesins, N. Livesey, A. M. Thompson, K. Walker, and J. C. Witte (2006), Seasonal cycles of O_3, CO, and convective outflow at the tropical tropopause, *Geophys. Res. Lett.*, *33*, L16,802, doi:10.1029/2006GL026602.

Froidevaux, L., et al. (2006), Early validation analyses of atmospheric profiles from EOS MLS on the Aura satellite, *IEEE Trans. Geosci. Remote Sens.*, *44*(5), 1106–1121.

Froidevaux, L., et al. (2007), Validation of EOS Microwave Limb Sounder stratospheric and mesospheric O_3 measurements, *This issue*.

Fu, R., et al. (2005), Short circuit of water vapor and polluted air to the global stratosphere by convective transport over the Tibetan Plateau, *Proc. Nat. Acad. Sci.*, *103*(5664–5669).

Intergovernmental Panel on Climate Change (2001), *Climate Change 2001, The Scientific Basis*, Cambridge University Press.

Jiang, Y. B., L. Froidevaux, et al. (2007), Comparison of EOS Microwave Limb Sounder O_3 data with ground-based and balloon observations, *This issue*.

Li, Q. B., et al. (2005), Convective outflow of South Asian pollution: A global CTM simulation compared with EOS MLS observations, *Geophys. Res. Lett.*, *32*(14), L14,826, doi:10.1029/2005GL022762.

Liu, H., D. J. Jacob, I. Bey, R. M. Yantosca, B. N. Duncan, and G. W. Sachse (2003), Transport pathways for Asian pollution outflow over the Pacific: Interannual and seasonal variations, *J. Geophys. Res.*, *108*(D20), 8786, doi:10.1029/2002JD003102.

Livesey, N. J., and W. G. Read (2000), Direct retrieval of line-of-sight atmospheric structure from limb sounding observations, *Geophys. Res. Lett.*, *27*(6), 891–894.

Livesey, N. J., W. V. Snyder, W. G. Read, and P. A. Wagner (2006), Retrieval algorithms for the EOS Microwave Limb Sounder (MLS), *IEEE Trans. Geosci. Remote Sens.*, *44*(5), 1144–1155.

Livesey, N. J., et al. (2005), EOS MLS version 1.5 Level 2 data quality and description document, *Tech. rep.*, Jet Propulsion Laboratory, D-32381.

Livesey, N. J., et al. (2007), EOS MLS version 2.2 Level 2 data quality and description document, *Tech. rep.*, Jet Propulsion Laboratory.

Marenco, A., et al. (1998), Measurement of ozone and water vapor by Airbus in-service aircraft: The MOZAIC airborne program, an overview, *J. Geophys. Res.*, *103*(D19), 25,631–25,642.

McGee, T. J., M. Gross, R. Ferrare, W. S. Heaps, and U. Singh (1993), Raman DIAL measurements of stratospheric zone in the presence of volcanic aerosols, *Geophys. Res. Lett.*, *20*, 955–958.

Nedelec, P., V. Thouret, J. Brioude, B. Sauvage, J.-P. Cammas, and A. Stohl (2005), Extreme CO concentrations in the upper troposphere over northeast Asia in June 2003 deom the in situ MOZAIC aircraft data, *Geophys. Res. Lett.*, *32*, L14,807.

Nédélec, P., et al. (2003), An improved infra-red carbon monoxide analyser for routine measurements aboard commercial Airbus aircraft: Technical validation and first scientific results of the MOZAIC III program, *Atmos. Chem. Phys.*, pp. 1551–1564.

Pearson, R. W., and D. H. Steadman (1980), Instrumentation for fast response ozone measurements from aircraft, *Atmos. Tech.*, *12*.

Prather, M. J., and D. J. Jacob (1997), A persistent imbalance in HO_x and NO_x photochemistry of the upper troposphere driven by deep tropical convection, *Geophys. Res. Lett.*, *24*(24), 3189–3192.

Pumphrey, H. C., et al. (2007), Validation of middle-atmosphere carbon monoxide retrievals from MLS on Aura, *Journal of Geophysical Research*, *xx*(xx), xx, (Submitted).

Read, W. G., Z. Shippony, M. J. Schwartz, N. J. Livesey, and W. V. Snyder (2006), The clear-sky unpolarized forward model for the EOS Microwave Limb Sounder (MLS), *IEEE Trans. Geosci. Remote Sens.*, *44*(5), 1367–1379.

Read, W. G., et al. (2007), EOS Aura Microwave Limb Sounder upper tropospheric and lower stratospheric humidity validation, in prep.

Rodgers, C. D. (2000), *Inverse Methods for Atmospheric Science, Theory and Practice*, 238 pp., World Scientific.

Santee, M. L., et al. (2007), Validation of Aura Microwave Limb Sounder nitric acid measurements, *This issue*.

Schoeberl, M. R., B. N. Duncan, A. R. Douglass, J. W. Waters, N. J. Livesey, W. G. Read, and M. J. Filipiak (2006a), The carbon monoxide tape recorder, *Geophys. Res. Lett.*, *33*, L12,811, doi: 10.1029/2006GL026178.

Schoeberl, M. R., et al. (2006b), Overview of the EOS Aura mission, *IEEE Trans. Geosci. Remote Sens.*, *44*(5), 1066–1074.

Schwartz, M. J., et al. (2007), Validation of the Aura Microwave Limb Sounder temperature and geopotential height measurements, *J. Geophys. Res.*, in preparation.

Stohl, A., S. Eckhardt, C. Forster, P. James, and N. Spichtinger (2002), On the pathways and timescales of intercontinental air pollution transport, *J. Geophys. Res.*, *107*(D23), 4684, doi:10.1029/2001JD001396.

Thouret, V., A. Marenco, J. A. Logan, P. Nédélec, and C. Grouhel (1998), Comparison of ozone measurements from the MOZAIC airborne programand the ozone sounding network at eight locations, *J. Geophys. Res.*, *103*, 25,695–25,720.

Waters, J. W., et al. (2006), The Earth Observing System Microwave Limb Sounder (EOS MLS) on the Aura satellite, *IEEE Trans. Geosci. Remote Sens.*, *44*(5), 1075–1092.

Webster, C. R., R. D. May, C. A. Trimble, R. G. Chave, and J. Kendall (1994), Aircraft (ER-2) Laser Infrared Absorption Spectrometer (ALIAS) for in situ stratospehric measurements of HCl, H_2O, CH_4, NO_2, and HNO_3, *Appl. Optics*, *33*, 454–472.

This preprint was prepared with AGU's LaTeX macros v5.01, with the extension package 'AGU^{++}' by P. W. Daly, version 1.6b from 1999/08/19.